Choosing the Number of Repetitions in the Multiple Plug-in Optimization Method for the Ridge Parameters in Multivariate Generalized Ridge Regression

(Last Modified: October 1, 2012)

Isamu Nagai *, Keisuke Fukui and Hirokazu Yanagihara

Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama,

Higashi-Hiroshima, Hiroshima 739-8526, Japan

Abstract

Multivariate generalized ridge (MGR) regression was proposed by Yanagihara, Nagai, and Satoh (2009) in order to avoid the multicollinearity problem in multivariate linear regression models. The MGR estimator is defined by using multiple nonnegative ridge parameters in an ordinary least-squares (LS) estimator. In order to optimize these ridge parameters, Yanagihara, Nagai, and Satoh (2009) and Nagai, Yanagihara, and Satoh (2012) proposed several optimization methods. We focus on the plug-in optimization method, which is an estimation method for the principal optimal ridge parameters that minimizes the predicted mean squared error. The plug-in optimization method is a repeating method that uses the current ridge parameters estimate as input in order to obtain an improved estimate. In the present paper, we propose two criteria for choosing the number of repetitions. We conducted several numerical studies using the proposed information criterion to compare the LS and MGR estimators with the optimized ridge parameters based on some ordinary plug-in optimization methods, and those obtained by using the optimized multiple plug-in optimization method.

Key words: Generalized ridge regression; Multivariate linear regression model; Plug-in optimization method; Shrinkage estimator.

1. Introduction

In the present paper, we consider a multivariate linear regression model with n observations of a p-dimensional vector of response variables and a k-dimensional vector of regressors (for more detailed information, see for example, Srivastava, 2002, Chapter 9; Timm, 2002, Chapter 4). Let $\boldsymbol{Y}$, $\boldsymbol{X}$, and $\boldsymbol{\mathcal{E}}$ be the $n \times p$ matrix of response variables, the $n \times k$ matrix of nonstochastic centered explanatory variables (i.e., $\boldsymbol{X}'\boldsymbol{1}_n = \boldsymbol{0}_k$) of rank$(\boldsymbol{X}) = k$, and the $n \times p$ matrix of error variables, respectively, where n is the sample size, $\boldsymbol{1}_n$ is an n-dimensional vector of ones, and $\boldsymbol{0}_k$ is a k-dimensional vector of zeros. Suppose that $n - k - p - 2 > 0$ and $\boldsymbol{\mathcal{E}} \sim N_{n \times p}(\boldsymbol{O}_{n \times p}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$, where $\boldsymbol{\Sigma}$ is a $p \times p$ unknown covariance matrix, rank$(\boldsymbol{\Sigma}) = p$, and $\boldsymbol{O}_{n \times p}$ is an $n \times p$ matrix of zeros. Then the matrix form of the

---

*Corresponding author, E-mail: inagai@hiroshima-u.ac.jp

multivariate linear regression model is expressed as

$$\boldsymbol{Y} = \boldsymbol{1}_n\boldsymbol{\mu}' + \boldsymbol{X}\boldsymbol{\Xi} + \boldsymbol{\mathcal{E}}, \tag{1.1}$$

where $\boldsymbol{\mu}$ is a p-dimensional unknown location vector and $\boldsymbol{\Xi}$ is a $k \times p$ unknown regression coefficient matrix. We can also express the model (1.1) as $\boldsymbol{Y} \sim N_{n \times p}(\boldsymbol{1}_n\boldsymbol{\mu}' + \boldsymbol{X}\boldsymbol{\Xi}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$.

The maximum-likelihood or least-squares (LS) estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Xi}$ are given by $\hat{\boldsymbol{\mu}} = \boldsymbol{Y}'\boldsymbol{1}_n/n$ and $\hat{\boldsymbol{\Xi}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, respectively, because $\boldsymbol{X}'\boldsymbol{1}_n = \boldsymbol{0}_k$. Since $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Xi}}$ are simple forms and are unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Xi}$, they are widely used in actual data analysis, see, e.g., Dien *et al.* (2006), Sârbu *et al.* (2008), Saxén and Sundell (2006), Skagerberg, Macgregor, and Kiparissides (1992), and Yoshimoto, Yanagihara, and Ninomiya (2005). However, when multicollinearity occurs in $\boldsymbol{X}$, the estimator of $\boldsymbol{\Xi}$ becomes unstable. In order to avoid this problem, multivariate generalized ridge (MGR) regression for the model in (1.1) was proposed by Yanagihara, Nagai, and Satoh (2009) (the original generalized ridge regression when $p = 1$ in the model (1.1) was proposed by Hoerl and Kennard (1970)). The MGR estimator is defined by using multiple ridge parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$, ($\theta_i \geq 0$, $i = 1, \ldots, k$). Nagai, Yanagihara, and Satoh (2012) showed that the principal optimal $\boldsymbol{\theta}$ that minimizes the predicted mean squared error (PMSE) is obtained in closed form with the unknown regression coefficient vector and covariance matrix.

In order to estimate the principal optimal $\boldsymbol{\theta}$, Nagai, Yanagihara, and Satoh (2012) proposed the plug-in optimization method. By replacing the LS and unbiased estimators for $\boldsymbol{\Sigma}$ with unknown values, the single plug-in optimized ridge parameters were derived. However, when multicollinearity occurs, the optimized ridge parameters tend to be too small since the LS estimator tends to have a large variance. Thus, to avoid under evaluation, Nagai, Yanagihara, and Satoh (2012) considered using the MGR estimator based on the single plug-in optimized ridge parameters instead of using the LS estimator. The double plug-in optimized ridge parameters were also derived. Repeating this renewal method, we obtained the multiple plug-in optimized ridge parameters derived from the MGR estimator based on the initial optimized ridge parameters. Let s ($s = 1, 2, 3, \ldots$) be the number of repetitions in the multiple plug-in optimization method. Note that the single plug-in optimized ridge parameters are obtained when $s = 1$ and the double ones are obtained when $s = 2$. In the present paper, we wish to find the value of s that minimizes the PMSE. In order to choose s, we propose two information criteria.

The remainder of the present paper is organized as follows: In Section 2, we illustrate the MGR estimator and a target PMSE. We also introduce the multiple plug-in optimization method. In Section 3, we propose some criteria for choosing the number of repetitions in the multiple plug-in optimization method. In Section 4, we compare the optimization methods by conducting numerical studies.

2. Preliminaries

In this section, we introduce the MGR estimator and the principal optimal $\boldsymbol{\theta}$ that minimizes the PMSE. Yanagihara, Nagai, and Satoh (2009) proposed the MGR estimator, which is defined as follows:

$$\hat{\boldsymbol{\Xi}}_{\boldsymbol{\theta}} = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{Q}\boldsymbol{\Theta}\boldsymbol{Q}')^{-1}\boldsymbol{X}'\boldsymbol{Y}, \tag{2.1}$$

where $\boldsymbol{\Theta} = \mathrm{diag}(\boldsymbol{\theta})$ is a $k \times k$ diagonal matrix and $\boldsymbol{Q}$ is a $k \times k$ orthogonal matrix that diagonalizes $\boldsymbol{X}'\boldsymbol{X}$, i.e., $\boldsymbol{Q}$ satisfies $\boldsymbol{Q}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{Q} = \boldsymbol{D}$, where $\boldsymbol{D} = \mathrm{diag}(d_1,\ldots,d_k)$ and $d_1,\ldots,d_k$ are the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$. This estimator coincides with the LS estimator when $\boldsymbol{\theta} = \boldsymbol{0}_k$, and it coincides with the ridge estimator for the model (1.1) proposed by Yanagihara and Satoh (2010) when $\boldsymbol{\theta} = \lambda\boldsymbol{1}_k$ for $\lambda \geq 0$.

Using the same orthogonal transform as was used by Nagai, Yanagihara, and Satoh (2012), the model (1.1) can be rewritten as follows:

$$\boldsymbol{Z} = \boldsymbol{L}\begin{pmatrix}\boldsymbol{\Gamma} \\ \boldsymbol{\mu}'\end{pmatrix} + \boldsymbol{\mathcal{V}},$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_k)' = \boldsymbol{Q}'\boldsymbol{\Xi}$, $\boldsymbol{Z} = (\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n)' = \boldsymbol{P}_1'\boldsymbol{Y}$, $\boldsymbol{\mathcal{V}} = (\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_n)' = \boldsymbol{P}_1'\boldsymbol{\mathcal{E}}$, $\boldsymbol{L} = (\mathrm{diag}(\sqrt{d_1},\ldots,\sqrt{d_k},\sqrt{n}), \boldsymbol{O}_{k+1,n-k-1})'$, and $\boldsymbol{P}_1$ is an $n \times n$ orthogonal matrix that diagonalizes $(\boldsymbol{X},\boldsymbol{1}_n)(\boldsymbol{X},\boldsymbol{1}_n)'$, that is, $\boldsymbol{P}_1$ satisfies $\boldsymbol{P}_1'(\boldsymbol{X},\boldsymbol{1}_n)(\boldsymbol{X},\boldsymbol{1}_n)'\boldsymbol{P}_1 = \boldsymbol{L}\boldsymbol{L}'$. When $p = 1$, this transformation was used in Goldstein and Smith (1974), and by others. Nagai, Yanagihara, and Satoh (2012) showed that $\boldsymbol{Z} \sim N_{n \times p}(\boldsymbol{\mathcal{M}}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_p)$, where $\boldsymbol{\mathcal{M}} = (\boldsymbol{m}_1,\ldots,\boldsymbol{m}_n)' = \boldsymbol{L}(\boldsymbol{\Gamma}',\boldsymbol{\mu})'$. The MGR estimator of $\boldsymbol{\Gamma}$ is defined by $\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\theta}} = \boldsymbol{Q}'\hat{\boldsymbol{\Xi}}_{\boldsymbol{\theta}}$, thus $\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\theta}} = (\boldsymbol{D}+\boldsymbol{\Theta})^{-1}\boldsymbol{C}'\boldsymbol{Z}$ where $\boldsymbol{C} = (\boldsymbol{D}^{1/2},\boldsymbol{O}_{k,n-k})'$, which is equivalent to the estimator obtained by substituting $\boldsymbol{D}+\boldsymbol{\Theta}$ into $\boldsymbol{D}$ in the LS estimator of $\boldsymbol{\Gamma}$, i.e., $\hat{\boldsymbol{\Gamma}} = \boldsymbol{D}^{-1}\boldsymbol{C}'\boldsymbol{Z}$.

Then the PMSE of $\hat{\boldsymbol{Z}}_{\boldsymbol{\theta}} = \boldsymbol{L}(\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\theta}}',\hat{\boldsymbol{\mu}})'$, which is the predictor of $\boldsymbol{Z}$, is defined as follows:

$$\mathrm{PMSE}[\hat{\boldsymbol{Z}}_{\boldsymbol{\theta}}] = E[r(\boldsymbol{V},\hat{\boldsymbol{Z}}_{\boldsymbol{\theta}})], \tag{2.2}$$

where $\boldsymbol{V} \sim N_{n \times p}(\boldsymbol{\mathcal{M}}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$, $\boldsymbol{V} \perp\!\!\!\perp \boldsymbol{Z}$, and the function $r(\cdot,\cdot)$ is defined by the following discrepancy function for measuring the distance between any $n \times p$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$r(\boldsymbol{A},\boldsymbol{B}) = \mathrm{tr}\{(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{\Sigma}^{-1}(\boldsymbol{A}-\boldsymbol{B})'\}. \tag{2.3}$$

From some simple calculations, we obtain $\mathrm{PMSE}[\hat{\boldsymbol{Z}}_{\boldsymbol{\theta}}] = np + E[r(\hat{\boldsymbol{Z}}_{\boldsymbol{\theta}},\boldsymbol{\mathcal{M}})]$. Nagai, Yanagihara, and Satoh (2012) showed that $\boldsymbol{\theta}^* = (\theta_1^*,\ldots,\theta_k^*)'$, the principal optimal $\boldsymbol{\theta}$ with minimized $E[r(\hat{\boldsymbol{Z}}_{\boldsymbol{\theta}},\boldsymbol{\mathcal{M}})]$, is derived as follows:

$$\theta_i^* = \frac{p}{\boldsymbol{\gamma}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}_i}, \quad (i = 1,\ldots,k). \tag{2.4}$$

3

However, θ_i^*, $(i = 1, \ldots, k)$ cannot be used for estimating $\boldsymbol{\Xi}$ since it includes the unknowns $\boldsymbol{\gamma}_i$ and $\boldsymbol{\Sigma}$.

Nagai, Yanagihara, and Satoh (2012) proposed the single plug-in optimization method by substituting $\hat{\boldsymbol{\gamma}}_i = \boldsymbol{z}_i/\sqrt{d_i}$, $(i = 1, \ldots, k)$, which is the ith row of $\hat{\boldsymbol{\Gamma}}$, for $\boldsymbol{\gamma}_i$, $(i = 1, \ldots, k)$; and $\boldsymbol{S} = \sum_{i=k+2}^{n} \boldsymbol{z}_i \boldsymbol{z}_i'/(n - k - 1)$, which is an unbiased estimator for $\boldsymbol{\Sigma}$, for $\boldsymbol{\Sigma}$ in $\hat{\theta}_i^*$, $(i = 1, \ldots, k)$ which is in equation (2.4). Then the estimator for θ_i^*, $(i = 1, \ldots, k)$ from the single plug-in optimization method is derived as follows:

$$\hat{\theta}_i^{[1]} = \frac{p}{\hat{\boldsymbol{\gamma}}_i' \boldsymbol{S}^{-1} \hat{\boldsymbol{\gamma}}_i} = \frac{d_i p}{t_i}, \quad (i = 1, \ldots, k), \tag{2.5}$$

where $t_i = \boldsymbol{z}_i' \boldsymbol{S}^{-1} \boldsymbol{z}_i$, $(i = 1, \ldots, k)$.

When multicollinearity occurs, we consider using the MGR estimator. However, in that case, $\hat{\theta}_i^{[1]}$, $(i = 1, \ldots, k)$ tends to be too small since it depends heavily on $\hat{\boldsymbol{\gamma}}_i$, and the variance of $\hat{\boldsymbol{\gamma}}_i$ becomes large. In order to avoid this problem, we can use the MGR estimator based on the single plug-in optimization method instead of using the LS estimator because the variance of the MGR estimator is smaller than that of the LS estimator. We then derive the double plug-in optimization method as

$$\hat{\theta}_i^{[2]} = \frac{p}{\hat{\boldsymbol{\gamma}}_i^{[1]'} \boldsymbol{S}^{-1} \hat{\boldsymbol{\gamma}}_i^{[1]}} = \left(1 + \frac{\hat{\theta}_i^{[1]}}{d_i}\right)^2 \hat{\theta}_i^{[1]} = \left(1 + \frac{p}{t_i}\right)^2 \frac{d_i p}{t_i}, \quad (i = 1, \ldots, k), \tag{2.6}$$

where $\hat{\boldsymbol{\gamma}}_i^{[1]} = \sqrt{d_i} \boldsymbol{z}_i/(d_i + \hat{\theta}_i^{[1]})$ is the ith row of $\hat{\boldsymbol{\Gamma}}_{\hat{\boldsymbol{\theta}}^{[1]}}$ obtained by substituting $\hat{\boldsymbol{\theta}}^{[1]} = (\hat{\theta}_1^{[1]}, \ldots, \hat{\theta}_k^{[1]})'$ for $\boldsymbol{\theta}$ in $\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\theta}}$. When we iterate this method, that is, we obtain new plug-in optimized ridge parameters by using the MGR estimator based on the current optimized ridge parameters, the multiple plug-in optimization method is derived. This was also proposed by Nagai, Yanagihara, and Satoh (2012), as follows:

$$\hat{\theta}_i^{[s]} = \frac{p}{\hat{\boldsymbol{\gamma}}_i^{[s-1]'} \boldsymbol{S}^{-1} \hat{\boldsymbol{\gamma}}_i^{[s-1]}} = \left(1 + \frac{\hat{\theta}_i^{[s-1]}}{d_i}\right)^2 \hat{\theta}_i^{[1]}, \quad (s = 1, 2, 3, \ldots; \ i = 1, \ldots, k), \tag{2.7}$$

where $\hat{\theta}_i^{[0]} = 0$ and $\hat{\boldsymbol{\gamma}}_i^{[s-1]}$ is the ith row of $\hat{\boldsymbol{\Gamma}}_{\hat{\boldsymbol{\theta}}^{[s-1]}}$; note that $\hat{\boldsymbol{\gamma}}_i^{[0]} = \hat{\boldsymbol{\gamma}}_i$, $(i = 1, \ldots, k)$. In the case of $p = 1$, $\hat{\theta}_i^{[1]}$ and $\hat{\theta}_i^{[2]}$ correspond with the optimization methods in Hoerl and Kennard (1970), and $\hat{\theta}_i^{[s]}$ coincides with the optimization method in Hemmerle (1975). Numerical studies in previous papers have compared only the single or double optimized ridge parameters. However, $\hat{\boldsymbol{\theta}}^{[s]} = (\hat{\theta}_1^{[s]}, \ldots, \hat{\theta}_k^{[s]})'$ is derived by using $\hat{\boldsymbol{\theta}}^{[s-1]}$ and (2.7) for any natural number s, and there is no method for choosing s. Hence we consider determining the value of s that reduces the PMSE of $\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}} = \boldsymbol{L}(\hat{\boldsymbol{\Gamma}}_{\hat{\boldsymbol{\theta}}^{[s]}}', \hat{\boldsymbol{\mu}})'$, which is the predictor of $\boldsymbol{Z}$ based on the multiple plug-in optimization method.

3. Method for Choosing s

In this section, we consider a method for choosing s, which was defined in (2.7) as the number of repetitions in the multiple plug-in optimization method. In order to choose s, our proposed

4

information criteria is to reduce the PMSE$[\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}]$. Note that $\hat{\theta}_i^{[s]}$ depends on t_i. In order to get right this dependence, we express $\hat{\theta}_i^{[s]}$ as $\hat{\theta}_i^{[s]}(t_i)$, $(s = 1, 2, \ldots; \ i = 1 \ldots, k)$. Letting $w^{[s]}(t_i) = d_i/(d_i + \hat{\theta}_i^{[s]}(t_i))$ for $i = 1, \ldots, k$, the ith row of $\hat{\boldsymbol{\Gamma}}_{\hat{\boldsymbol{\theta}}^{[s]}}$ is obtained by $\hat{\boldsymbol{\gamma}}_i^{[s]} = w^{[s]}(t_i)\hat{\boldsymbol{\gamma}}_i$. The s that minimizes PMSE$[\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}]$ is regarded as the optimal s, and we now consider how to estimate it. Hence we consider evaluating $\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}$, which is obtained by using $\hat{\boldsymbol{\theta}}^{[s]} = (\hat{\theta}_1^{[s]}(t_1), \ldots, \hat{\theta}_k^{[s]}(t_k))'$, by stating the PMSE as follows:

$$\text{PMSE}[\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}] = E[r(\boldsymbol{V}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}})].$$

The predicted value $\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}$ can be expressed as $\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}} = (\hat{\boldsymbol{z}}_1(\hat{\theta}_1^{[s]}(t_1)), \ldots, \hat{\boldsymbol{z}}_k(\hat{\theta}_k^{[s]}(t_k)), \hat{\boldsymbol{z}}_{k+1}, \ldots, \hat{\boldsymbol{z}}_n)'$ since the ith row vector of $\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}$ depends on $\hat{\theta}_i^{[s]}(t_i)$ for $i = 1, \ldots, k$, and it does not depend on $\hat{\boldsymbol{\theta}}^{[s]}$ for $i = k+1, \ldots, n$. Additionally, it should be kept in mind that $\hat{\boldsymbol{z}}_i(\hat{\theta}_i^{[s]}(t_i)) = w^{[s]}(t_i)\boldsymbol{z}_i$, $(i = 1, \ldots, k)$, $\hat{\boldsymbol{z}}_{k+1} = \boldsymbol{z}_{k+1}$, $\hat{\boldsymbol{z}}_i = \boldsymbol{0}_p$, $(i = k+2, \ldots, n)$, and $\boldsymbol{z}_i \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{m}_i, \boldsymbol{\Sigma})$ where $\boldsymbol{m}_i$ is the ith row of $\boldsymbol{\mathcal{M}} = \boldsymbol{L}(\boldsymbol{\Gamma}', \boldsymbol{\mu})'$. From a simple calculation, we obtain

$$\text{PMSE}[\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}] = E[r(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}})] + 2E[\text{tr}\{(\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}} - \boldsymbol{\mathcal{M}})'(\boldsymbol{Z} - \boldsymbol{\mathcal{M}})\boldsymbol{\Sigma}^{-1}\}]. \tag{3.1}$$

Then we can see that

$$E[\text{tr}\{(\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}} - \boldsymbol{\mathcal{M}})'(\boldsymbol{Z} - \boldsymbol{\mathcal{M}})\boldsymbol{\Sigma}^{-1}\}]$$
$$= \sum_{i=1}^{k} E\left[\left(w^{[s]}(t_i)\boldsymbol{z}_i - \boldsymbol{m}_i\right)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{m}_i)\right] + E[(\boldsymbol{z}_{k+1} - \boldsymbol{m}_{k+1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_{k+1} - \boldsymbol{m}_{k+1})]$$
$$+ \sum_{i=k+2}^{n} E[\boldsymbol{m}_i'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{m}_i)]$$
$$= \sum_{i=1}^{k} E[w^{[s]}(t_i)\boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{m}_i)] + p,$$

because $E[(\boldsymbol{z}_{k+1} - \boldsymbol{m}_{k+1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_{k+1} - \boldsymbol{m}_{k+1})] = \text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}) = p$ and $E[\boldsymbol{m}_i'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{m}_i)] = 0$. If we let $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ip})' = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{z}_i - \boldsymbol{m}_i)$, where $\boldsymbol{\Sigma}^{1/2}$ satisfies $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$, then $\boldsymbol{u}_i \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{0}_p, \boldsymbol{I}_p)$ and the following result is derived:

$$E[\text{tr}\{(\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}} - \boldsymbol{\mathcal{M}})'(\boldsymbol{Z} - \boldsymbol{\mathcal{M}})\boldsymbol{\Sigma}^{-1}\}] = \sum_{i=1}^{k} E[w^{[s]}(t_i)(\boldsymbol{u}_i + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{m}_i)'\boldsymbol{u}_i] + p$$
$$= \sum_{i=1}^{k}\sum_{j=1}^{p} E[w^{[s]}(t_i)(u_{ij} + \boldsymbol{m}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{e}_{p\cdot j})u_{ij}] + p,$$

where $\boldsymbol{e}_{p\cdot j}$ is a p-dimensional vector in which only the jth element is equal to one and the other

elements are zeros. Using the formula in Stein (1981) (see, e.g., Efron (2004)), we obtain

$$E[w^{[s]}(t_i)(u_{ij} + \boldsymbol{m}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{e}_{p\cdot j})u_{ij}] = E\left[\frac{\partial\ \{w^{[s]}(t_i)(u_{ij} + \boldsymbol{m}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{e}_{p\cdot j})\}}{\partial u_{ij}}\right]$$

$$= E\left[\frac{\partial w^{[s]}(t_i)}{\partial u_{ij}}(u_{ij} + \boldsymbol{m}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{e}_{p\cdot j}) + w^{[s]}(t_i)\right]$$

$$= E\left[\boldsymbol{e}_{p\cdot j}'\frac{\partial w^{[s]}(t_i)}{\partial \boldsymbol{u}_i}\boldsymbol{e}_{n\cdot i}'(\boldsymbol{U} + \boldsymbol{\mathcal{M}}\boldsymbol{\Sigma}^{-1/2})\boldsymbol{e}_{p\cdot j} + w^{[s]}(t_i)\right],$$

where $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)'$. Since $\boldsymbol{U} + \boldsymbol{\mathcal{M}}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{Z}\boldsymbol{\Sigma}^{-1/2}$ and t_i depends on $\boldsymbol{u}_i$, the following equation is derived:

$$E[w^{[s]}(t_i)(u_{ij} + \boldsymbol{m}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{e}_{p\cdot j})u_{ij}] = E\left[\boldsymbol{e}_{p\cdot j}'\frac{\partial t_i}{\partial \boldsymbol{u}_i}\frac{\partial w^{[s]}(t_i)}{\partial t_i}\boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{e}_{p\cdot j} + w^{[s]}(t_i)\right],$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, p$. Note that $t_i = \boldsymbol{z}_i'\boldsymbol{S}^{-1}\boldsymbol{z}_i = (\boldsymbol{\Sigma}^{1/2}\boldsymbol{u}_i + \boldsymbol{m}_i)'\boldsymbol{S}^{-1}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{u}_i + \boldsymbol{m}_i)$ and that $\boldsymbol{S}^{-1}$ does not depend on $\boldsymbol{u}_i$, $(i = 1, \ldots, k)$ since $\boldsymbol{S}$ does not depend on $\boldsymbol{z}_i$, $(i = 1, \ldots, k+1)$ and $\boldsymbol{u}_i$ is obtained from $\boldsymbol{z}_i$. Thus we obtain the following differential result:

$$\frac{\partial t_i}{\partial \boldsymbol{u}_i} = 2\boldsymbol{\Sigma}^{1/2}\boldsymbol{S}^{-1}\boldsymbol{z}_i.$$

Hence we calculate (3.1) as follows:

$$\text{PMSE}[\hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}] = E[r(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}})] + 2p + 2\sum_{i=1}^{k} E\left[\boldsymbol{z}_i'\boldsymbol{\Sigma}^{-1/2}\frac{\partial t_i}{\partial \boldsymbol{u}_i}\frac{\partial w^{[s]}(t_i)}{\partial t_i} + pw^{[s]}(t_i)\right]$$

$$= E[r(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}})] + 2p + 2\sum_{i=1}^{k} E\left[2t_i\frac{\partial w^{[s]}(t_i)}{\partial t_i} + pw^{[s]}(t_i)\right].$$

From this result, the unbiased estimator of (3.1) can be defined as follows:

$$C_p^* = r(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}) + 2p + 2\sum_{i=1}^{k}\left(2t_i\frac{\partial w^{[s]}(t_i)}{\partial t_i} + pw^{[s]}(t_i)\right). \tag{3.2}$$

However, we cannot use this criterion since $r(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}})$ requires the unknown covariance matrix $\boldsymbol{\Sigma}$.

In (3.2), we will consider estimating C_p^* by using the idea for the C_p and MC_p criteria that was put forth in Yanagihara, Nagai, and Satoh (2009). We thus estimate the C_p^* criterion as follows:

Definition 1. *The criteria for choosing s are defined by*

$$C_p^\# = \hat{r}(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}) + 2p + 2\sum_{i=1}^{k}\left(2t_i\frac{\partial w^{[s]}(t_i)}{\partial t_i} + pw^{[s]}(t_i)\right), \tag{3.3}$$

$$MC_p^\# = c_M\hat{r}(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}}) + 2p + 2\sum_{i=1}^{k}\left(2t_i\frac{\partial w^{[s]}(t_i)}{\partial t_i} + pw^{[s]}(t_i)\right) + p(p+1), \tag{3.4}$$

where $c_M = 1 - (p+1)/(n-k-1)$, and $\hat{r}(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}})$ is obtained by substituting $\boldsymbol{S}$ for $\boldsymbol{\Sigma}$ in $r(\boldsymbol{Z}, \hat{\boldsymbol{Z}}_{\hat{\boldsymbol{\theta}}^{[s]}})$.

Minimizing each criterion, we can obtain several estimators for the optimal s. Let s_{C}^* and s_{M}^* be obtained by minimizing the $C_p^\#$ and $MC_p^\#$ criteria, respectively. From (2.7), we can obtain $\hat{\boldsymbol{\theta}}^{[s_{\mathrm{C}}^*]} = (\hat{\theta}_1^{[s_{\mathrm{C}}^*]}, \ldots, \hat{\theta}_k^{[s_{\mathrm{C}}^*]})'$ and $\hat{\boldsymbol{\theta}}^{[s_{\mathrm{M}}^*]} = (\hat{\theta}_1^{[s_{\mathrm{M}}^*]}, \ldots, \hat{\theta}_k^{[s_{\mathrm{M}}^*]})'$ in closed forms, respectively. Then, we can obtain $\hat{\boldsymbol{\Xi}}_{\hat{\boldsymbol{\theta}}^{[s_{\mathrm{C}}^*]}}$ and $\hat{\boldsymbol{\Xi}}_{\hat{\boldsymbol{\theta}}^{[s_{\mathrm{M}}^*]}}$ by substituting $\hat{\boldsymbol{\theta}}^{[s_{\mathrm{C}}^*]}$ and $\hat{\boldsymbol{\theta}}^{[s_{\mathrm{M}}^*]}$ into (2.1), respectively.

4. Numerical Studies

By conducting numerical studies, we compared the PMSEs of the predictors of $\boldsymbol{Y}$ consisting of the ridge regression estimators with the optimized ridge parameters by using each method. Let $\boldsymbol{R}_q = \mathrm{diag}(\sqrt{1}, \ldots, \sqrt{q})$, which is a $q \times q$ diagonal matrix, and let $\boldsymbol{\Delta}_q(\rho)$ be a $q \times q$ matrix whose (i, j)th element is $\rho^{|i-j|}$. Then the explanatory matrix is $\boldsymbol{X} = \boldsymbol{W}\boldsymbol{\Psi}^{1/2}$, where $\boldsymbol{\Psi} = \boldsymbol{R}_k\boldsymbol{\Delta}_k(\rho_x)\boldsymbol{R}_k$, and $\boldsymbol{W}$ is an $n \times k$ matrix whose elements were generated independently from the uniform distribution on $(-1, 1)$. The $k \times p$ unknown regression coefficient matrix $\boldsymbol{\Xi}$ is defined by $\boldsymbol{\Xi} = \delta\boldsymbol{F}\boldsymbol{\Pi}$, where δ is a constant term, $\boldsymbol{F}$ is defined as $\boldsymbol{F} = \mathrm{diag}(\boldsymbol{1}_\kappa, \boldsymbol{0}_{k-\kappa})$, which is a $k \times k$ diagonal matrix, and $\boldsymbol{\Pi}$ is defined by $\boldsymbol{1}'_{p/3} \otimes \boldsymbol{\Pi}_1$ when $k = 10$ and by $\boldsymbol{1}'_{p/3} \otimes \boldsymbol{\Pi}_2$ when $k = 15$. Here, $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ are given by

$$
\boldsymbol{\Pi}_1 = \begin{pmatrix}
0.8501 & 0.6571 & 0.2159 \\
-0.2753 & -0.2432 & -0.1187 \\
-0.3193 & -0.2926 & -0.1671 \\
0.2754 & 0.2608 & 0.1766 \\
0.2693 & 0.2164 & 0.2066 \\
-0.0676 & -0.0663 & -0.0561 \\
0.2239 & 0.2197 & 0.1880 \\
-0.0352 & -0.0346 & -0.0305 \\
0.3240 & 0.3199 & 0.2868 \\
-0.3747 & -0.3727 & -0.3554
\end{pmatrix}, \quad
\boldsymbol{\Pi}_2 = \begin{pmatrix}
1.3794 & 0.0645 & 0.0330 \\
-0.0766 & -0.0241 & -0.0143 \\
-0.2618 & -0.1396 & -0.0951 \\
-0.4619 & -0.2589 & -0.1798 \\
0.2381 & 0.1488 & 0.1082 \\
0.2140 & 0.1463 & 0.1112 \\
0.3002 & 0.2364 & 0.1950 \\
0.1155 & 0.0953 & 0.0812 \\
-0.2774 & -0.2395 & -0.2091 \\
0.3392 & 0.3072 & 0.2807 \\
0.0016 & 0.0107 & 0.0100 \\
0.0438 & 0.0408 & 0.0381 \\
-0.3187 & -0.3039 & -0.2904 \\
0.0529 & 0.0510 & 0.0493 \\
0.2505 & 0.2451 & 0.2399
\end{pmatrix}.
$$

Here, δ controls the scale of the regression coefficient matrix, and $\boldsymbol{F}$ controls the number of nonzero regression coefficients via κ. The values of the elements of $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$, which are an essential regression coefficient matrix, are the same as in Lawless (1981). The simulated data $\boldsymbol{Y}$ were generated iteratively from $N_{n\times p}(\boldsymbol{X}\boldsymbol{\Xi}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$ under several selections of n, k, κ, δ, ρ_y, and ρ_x, where $\boldsymbol{\Sigma} = \boldsymbol{R}_p\boldsymbol{\Delta}_p(\rho_y)\boldsymbol{R}_p$, and the number of iterations was $10,000$. At each iteration, we evaluated $r(\boldsymbol{X}\boldsymbol{\Xi}, \hat{\boldsymbol{Y}}_{\hat{\boldsymbol{\theta}}})$ where $\hat{\boldsymbol{Y}}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{1}_n\hat{\boldsymbol{\mu}}' + \boldsymbol{X}\hat{\boldsymbol{\Xi}}_{\hat{\boldsymbol{\theta}}}$, which is the predicted value of $\boldsymbol{Y}$ obtained from each method. The average of $np + r(\boldsymbol{X}\boldsymbol{\Xi}, \hat{\boldsymbol{Y}}_{\hat{\boldsymbol{\theta}}})$ across $10,000$ iterations was regarded as the PMSE of $\hat{\boldsymbol{Y}}_{\hat{\boldsymbol{\theta}}}$. In the simulation, a standardized $\boldsymbol{X}$ was used to estimate the regression coefficients.

We obtained the optimized ridge parameter $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)'$ from the following two methods:

Method 1 : $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{[s_C^*]}$ where $s_C^* = \arg\min_{s \in \mathcal{S}} C_p^\#$ and $C_p^\#$ is defined in (3.3).

Method 2 : $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{[s_M^*]}$ where $s_M^* = \arg\min_{s \in \mathcal{S}} MC_p^\#$ and $MC_p^\#$ is defined in (3.4).

In this paper, we set $\mathcal{S} = \{1, 2, 3, 4, 5, 10, 15, 20, 50\}$ and let $\chi(i) = i$ for $i = 1, \ldots, 5$, $\chi(i) = 5(i - 4)$ for $i = 6, 7, 8$, and $\chi(9) = 50$, and we let $\sharp(\mathcal{S})$ be the number of elements in the set $\mathcal{S}$. To reduce the number of computations, we applied the selection method proposed by Ruppert (2002) to select $s \in \mathcal{S}$, as follows:

1. Set i to 1.

2. Calculate several information criteria (IC) for $\chi(i)$ and $\chi(i+1)$, and denote these IC values as IC(i) and IC($i + 1$), respectively.

3. If IC($i + 1$) > $0.98 \times$ IC(i), stop iterating and go to Step 5.

4. If $i + 1 \leq \sharp(\mathcal{S})$, update i as $i + 1$ and go to Step 2. Otherwise, if $i + 1 > \sharp(\mathcal{S})$, do not update i and go to Step 5.

5. When IC($i + 1$) < IC(i), let $i^* = i + 1$; otherwise, let $i^* = i$.

6. Obtain the optimized s as $\chi(i^*)$.

By using this selection method, we can reduce the number of computations for selecting s since it stops when there is little improvement in the information criteria when i is large, i.e., when s becomes large. When we use the $C_p^\#$ criterion to obtain s, we calculate (3.3) to obtain IC(i). As was done for $C_p^\#$, $MC_p^\#$ in (3.4) is calculated for each $s \in \mathcal{S}$. For the purpose of comparison with the proposed methods, we prepared the two conventional optimization methods, as follows:

Method 3 : $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{[1]} = (\hat{\theta}_1^{[1]}, \ldots, \hat{\theta}_k^{[1]})'$, which is the single plug-in optimization method in (2.5).

Method 4 : $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{[2]} = (\hat{\theta}_1^{[2]}, \ldots, \hat{\theta}_k^{[2]})'$, which is the double plug-in optimization method in (2.6).

Table 1 shows the simulation results for $\mathrm{PMSE}[\hat{\boldsymbol{Y}}_{\hat{\boldsymbol{\theta}}}]/\{p(n + k + 1)\} \times 100$ for the cases in which $(k, n) = (10, 30)$ and $(10, 50)$, and Table 2 shows the results for the cases in which $(15, 30)$ and $(15, 50)$. In both tables, $p = 6$, where $p(n + k + 1)$ is the theoretical value of the PMSE for the predictor of $\boldsymbol{Y}$ based on the LS estimators.

We can see that all of the methods improved the PMSEs of the LS estimators in all cases since none of the values in the tables exceed 100. When $k = 10$, Method 2 is almost always the best method for small δ and n. Methods 1 and 2 resulted in a greater improvement than did Method 3 in almost all cases when $k = 10$ and $n = 30$. Methods 1 and 2 resulted in a greater improvement than did Method 3 in almost all cases when $k = 10$, $n = 50$, and δ was small. When $k = 15$,

Table 1. The results when $k = 10$

κ	δ	ρ_y	ρ_x	$n = 30$ Method				$n = 50$ Method			
				1	2	3	4	1	2	3	4
0	0	0.2	0.2	*79.71*	**78.25**	85.46	80.92	*85.28*	**84.82**	89.30	86.10
			0.95	*79.75*	**78.29**	85.50	80.96	*85.31*	**84.85**	89.32	86.12
		0.95	0.2	*79.76*	**78.31**	85.50	80.97	*85.33*	**84.86**	89.34	86.13
			0.95	*79.73*	**78.28**	85.49	80.94	*85.35*	**84.88**	89.36	86.16
5	1	0.2	0.2	86.47	**86.32**	88.98	*86.46*	91.45	91.54	92.73	**91.39**
			0.95	*82.28*	**81.16**	86.88	83.02	*87.50*	**87.09**	90.58	87.93
		0.95	0.2	*90.29*	90.63	91.62	**90.21**	*94.20*	94.29	94.81	**94.19**
			0.95	*83.35*	**82.24**	87.58	83.97	*88.72*	**88.39**	91.27	88.94
	3	0.2	0.2	*93.63*	93.90	94.31	**93.59**	97.33	97.24	**97.13**	*97.23*
			0.95	*85.44*	**84.72**	88.69	85.68	90.62	**90.52**	92.36	*90.60*
		0.95	0.2	*96.20*	*96.20*	96.36	**96.08**	98.23	98.24	**98.09**	*98.12*
			0.95	*89.00*	89.04	90.88	**88.95**	*92.93*	93.00	94.00	**92.90**
10	1	0.2	0.2	*87.12*	**86.67**	89.75	87.21	*94.19*	94.29	94.67	**94.17**
			0.95	*83.75*	**82.77**	87.69	84.22	*88.83*	**88.51**	91.29	89.01
		0.95	0.2	*91.09*	91.43	92.25	**91.03**	96.47	96.45	**96.30**	*96.44*
			0.95	*84.65*	**83.76**	88.27	85.02	*89.68*	**89.46**	91.79	89.75
	3	0.2	0.2	*93.53*	93.95	94.14	**93.49**	*98.76*	98.87	**98.72**	98.96
			0.95	*89.76*	89.91	91.37	**89.69**	*93.26*	93.35	94.21	**93.23**
		0.95	0.2	98.12	98.09	**97.74**	*98.05*	**99.51**	**99.51**	**99.51**	*99.67*
			0.95	*91.94*	92.36	92.88	**91.89**	*94.78*	94.84	95.29	**94.77**
	Average			*87.28*	**86.81**	90.07	87.62	*91.88*	**91.75**	93.50	92.09

Method 2 is always the best method for small δ and large ρ_x. Methods 1 and 2 resulted in a greater improvement than did Method 3 in almost cases when $k = 15$. Methods 1 and 2 also resulted in a greater improvement than did Method 4 in all cases when δ was small and ρ_x was large. When δ was small and ρ_x was large, Method 2 also resulted in a greater improvement than did Method 1 in almost all cases. Methods 1 and 2 resulted in the greatest improvement when k became large and ρ_y was small. In almost all cases, there was greater improvement when κ was smaller . When n or δ became small, each method was improved. On average, Method 2 was the best method, and Method 1 was the second best. Hence, we recommend using the $MC_p^{\#}$ criterion in (3.4) to choose the number of repetitions in the multiple plug-in optimization method.

References

[1] Dien, S. J. V., Iwatani, S. Usuda, Y. & Matsui, K. (2006). Theoretical analysis of amino acid-producing *Eschenrichia coli* using a stoixhiometrix model and multivariate linear regression, *J. Biosci. Bioeng.*, **102**, 34–40.

[2] Efron, B. (2004). The estimation of prediction error: covariance penalties and cross-validation. *J. Amer. Statist. Assoc.*, **99**, 619–632

Table 2. The results when $k = 15$

| | | | | $n = 30$ | | | | $n = 50$ | | | |
| | | | | Method | | | | Method | | | |
κ	δ	ρ_y	ρ_x	1	2	3	4	1	2	3	4
0	0	0.2	0.2	*74.89*	**72.16**	82.13	76.43	*79.46*	**78.83**	85.39	80.96
			0.95	*74.90*	**72.18**	82.10	76.42	*79.42*	**78.81**	85.36	80.94
		0.95	0.2	*74.87*	**72.15**	82.09	76.40	*79.47*	**78.84**	85.42	80.98
			0.95	*74.91*	**72.20**	82.13	76.43	*79.43*	**78.82**	85.37	80.95
5	1	0.2	0.2	*81.80*	**81.09**	85.64	81.99	*86.56*	**86.40**	88.98	86.57
			0.95	*77.29*	**75.13**	83.28	78.26	*82.00*	**81.54**	86.56	82.85
		0.95	0.2	*90.51*	90.89	91.80	**90.47**	*94.27*	94.36	94.61	**94.27**
			0.95	*84.04*	**83.56**	87.26	84.14	*87.74*	**87.55**	89.98	*87.74*
	3	0.2	0.2	*90.43*	90.79	91.75	**90.40**	**93.84**	*93.95*	94.35	**93.84**
			0.95	*83.64*	**83.21**	86.90	83.72	*87.39*	**87.19**	89.68	87.40
		0.95	0.2	*96.06*	*96.06*	96.32	**95.91**	98.10	98.05	*98.03*	**97.90**
			0.95	90.43	**90.39**	92.03	*90.40*	*92.12*	92.17	93.36	**92.08**
10	1	0.2	0.2	*84.73*	84.93	87.37	**84.68**	*90.53*	90.91	91.49	**90.45**
			0.95	*78.56*	**76.48**	84.17	79.44	*83.20*	**82.71**	87.43	83.95
		0.95	0.2	*93.55*	94.03	93.97	**93.52**	96.50	*96.49*	**96.19**	*96.49*
			0.95	*83.57*	**82.79**	87.06	83.74	*87.64*	**87.40**	89.98	87.69
	3	0.2	0.2	*94.13*	94.48	94.57	**94.10**	98.00	97.99	**97.65**	*97.96*
			0.95	*85.13*	**84.92**	87.94	85.14	*89.86*	90.01	91.29	**89.77**
		0.95	0.2	98.60	98.73	**98.48**	98.68	*99.34*	*99.34*	**99.33**	99.46
			0.95	*91.32*	91.67	92.51	**91.29**	**94.01**	*94.12*	94.58	**94.01**
15	1	0.2	0.2	*88.25*	89.06	89.86	**88.19**	*92.51*	92.76	93.08	**92.49**
			0.95	*79.46*	**77.67**	84.59	80.16	*84.13*	**83.72**	87.88	84.66
		0.95	0.2	*93.14*	93.53	93.81	**93.12**	96.90	*96.86*	**96.70**	*96.86*
			0.95	*84.09*	**83.47**	87.38	84.22	*88.20*	**88.01**	90.36	88.21
	3	0.2	0.2	96.43	96.46	*96.39*	**96.31**	*98.47*	98.68	**98.29**	98.66
			0.95	86.45	**86.28**	88.91	*86.44*	*91.05*	91.28	92.16	**90.99**
		0.95	0.2	*98.61*	98.76	**98.49**	98.71	**99.36**	**99.36**	**99.36**	*99.48*
			0.95	*90.91*	91.03	92.34	**90.88**	*94.13*	94.22	94.70	**94.12**
	Average			*86.45*	**85.86**	89.33	86.77	*90.13*	**90.01**	92.06	90.42

[3] Goldstein, M. & Smith, A. F. M. (1974), Ridge-type estimators for regression analysis, *J. Roy. Statist. Soc. Ser. B*, **36**, 284–291.

[4] Hemmerle, W. J. (1975), An explicit solution for generalized ridge regression, *Technometrics*, **17**, 309–314.

[5] Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.

[6] Lawless, J. F. (1981). Mean squared error properties of generalized ridge estimators, *J. Amer. Statist. Assoc.*, **76**, 462–466.

[7] Nagai, I., Yanagihara, H. & Satoh, K. (2012). Optimization of ridge parameters in multivariate generalized ridge regression by plug-in methods. *Hiroshima Math. J.* (in press).

[8] Ruppert, D. (2002). Selection the number of knots for penalized splines, *J. Comput. Graph. Statist.*, **11**, 735–757.

[9] Sârbu, C., Onisor, C., Posa, Mihalj, Kevresan, S. & Kuhajda, K. (2008). Modeling and prediction (correction) of partition coefficients of bile acids and their derivatives by multivariate regression methods, *Talanta*, **75**, 651–657.

[10] Saxén, R. & Sundell, J. (2006). [137]Cs in freshwater fish in Finland since 1986– a statistical analysis with multivariate linear regression models, *J. Environ. Radioactiv.*, **87**, 62–76.

[11] Skagerberg, B. , MacGregor, J. & Kiparissides, C. (1992). Multivariate data analysis applied to low-density polyethylene reactors, *Chemometr. Intell. Lab. Syst.*, **14**, 341–356.

[12] Srivastava, M. S. (2002). *Methods of Multivariate Statistics*, John Wiley & Sons, New York.

[13] Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution, *Ann. Statist.*, **9**, 1135–1151.

[14] Timm, N. H. (2002). *Applied Multivariate Analysis*, Springer-Verlag, New York.

[15] Yanagihara, H., Nagai, I. & Satoh, K. (2009). A bias-corected C_p criterion for optimizing ridge parameters in multivariate generalized ridge regression. *Japanese J. Appl. Statist.*, **38**, 151-172 (in Japanese).

[16] Yanagihara, H. & Satoh, K. (2010). An unbiased C_p criterion for multivariate ridge regression, *J. Multivariate Anal.*, **101**, 1226–1238.

[17] Yoshimoto, A., Yanagihara, H. & Ninomiya, Y. (2005). Finding factors affecting a forest stand growth through multivariate linear modeling, *J. Jpn. For. Res.*, **87**, 504–512 (in Japanese).